VOICEXML-BASED SPOKEN LANGUAGE INTERACTIVE SYSTEM

Ing. Stanislav Ondáš
supervisor: doc. Ing. Jozef Juhár, CSc.

Department of Electronics and Multimedia Communications,
Technical University of Košice, Park Komenského 13, 041 20 Košice,
Slovak Republic.
Tel: +421-95-6338659; Fax: +421-95-6323989
E-mail: Stanislav.Ondas@tuke.sk

ABSTRACT

The paper deals with architecture of spoken language interactive system (SLIS) and implementation of VoiceXML based dialog manager to SLIS. There are two basic architectures commonly used for SLIS – pipeline and distributed. Distributed architecture reflects growing requirements on dialog interaction and provides more flexibility and modularity. The Galaxy communicator based on Hub architecture is the typical example of distributed system. Our spoken language system is built on Hub architecture and implements VoiceXML based dialog manager, which handles tasks of natural language understanding (NLU) and natural language generation (NLG). The weather information and traintable information services in Slovak are available on our SLIS at this time.

1 INTRODUCTION

The verbal communication is the most natural form of interaction for people. So the human speech is very good medium for telecommunication services accessed by phone. There is an effort to use this kind of communication also in human – computer interaction and for the data accessing from world-wide network – Internet.

The systems that enable communication with computer or information gaining by voice are called Spoken Language Interaction Systems (SLIS).

In the next parts are discussed SLIS architectures with their basic components and implementation of VoiceXML based dialog manager to the Galaxy architecture.

2 THE GENERAL MODEL OF SLIS

The general block diagram of spoken language interaction system is shown on the Figure 1. It consists of the following basic components:

Input – Output block ***(I/O)*** is responsible for preprocessing and postprocessing of speech signal.
Automatic Speech Recognition ***(ASR)*** performs speech to text conversion.
Natural Language Understanding ***(NLU)*** extracts meaning of the recognized words and phrases in application's context.
Dialog Manager ***(DM)*** manages all system operations. It generates appropriate system's response based on input information flow, actual dialogue state and its history.

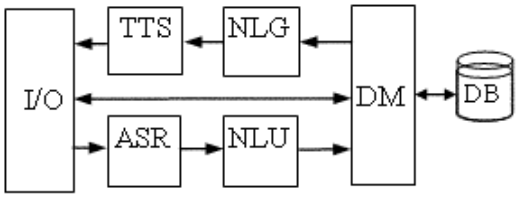

Fig.1: The general model of SLIS

Natural Language Generation ***(NLG)*** is responsible for representation of information returned back to the user.
Text to Speech ***(TTS)*** generates phonetic representation and speech signal forming system's response.

3 ARCHITECTURES OF SLIS

From the processing flow and controlling point of view we can divide architectures of spoken language systems into two categories:

- Pipeline architectures
- Distributed architectures

3.1 Pipeline architectures

Pipeline architecture is the oldest type of architecture and it is derived from the general model of SLIS, as shown on Figure 1. All system's components work in strictly specified order.

Low flexibility and late responses to errors are the main handicaps of pipeline architecture.

3.2 Distributed architectures

Transition from the pipeline approach to a distributed architecture reflects the growing requirements on spoken language systems.

Distributed architectures are more modular. They provide ability for process management, and make easier the way for adding new features.

A typical example of distributed approach is Hub architecture as used in Galaxy communicator.

Hub architecture was developed in the project supported by the grant agency DARPA. DARPA communicator [2] is an open system, where the main component is called "Hub". Hub process distributes the communication and services between other components - servers. The successor of DARPA communicator is The Galaxy communicator [2]. Its scheme is shown on Figure 2. The Galaxy Communicator is distributed, message-based, hub-and-spoke software infrastructure optimized for spoken dialogue systems development.

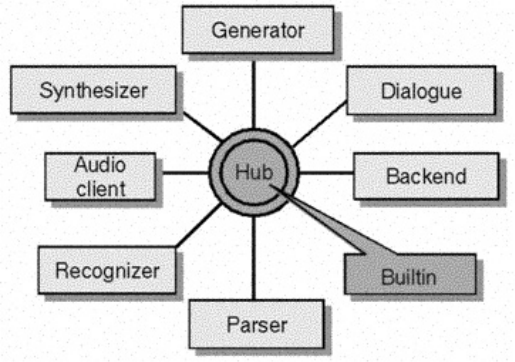

Fig. 2: Galaxy communicator

The components as Generator (NLG), Synthesizer (TTS), Recognizer (ASR), Dialogue (DM), and Parser (NLU) were described earlier.

Audio server provides voice data and telephony events to the system and is the same as I/O block mentioned earlier.

Backend server is responsible for communication with external data sources and the **Builtin** server provides some build in features to the system. These features are the part of Galaxy software distribution.

4 GALAXY ARCHITECTURE WITH VOICEXML BASED DM

In the research project Intelligent Speech Communication Interface (IRKR) we are building an own SLIS based on Galaxy architecture. This architecture combines flexibility and modularity of Hub architecture and the powerful tool for voice dialogs – the VoiceXML language.

Dialog Manager (DM) is implemented as the VoiceXML language interpreter. Architecture of the IRKR communicator is shown on the Figure 3.

The key change compared to the original Galaxy infrastructure is the merger of NLG (Generator), NLU (Parser) and DM (Dialog) servers. The new arose server is also called Dialog Manager (DM).

This VoiceXML based DM is able to handle all tasks related to above mentioned processes.

The VoiceXML language with other W3C Speech Interface Framework (www.w3c.org) languages are ready to manage the tasks of user input stream parsing and its semantic interpretation, generation of dialog states, recording of dialog context and system output generation in structural and textual form.

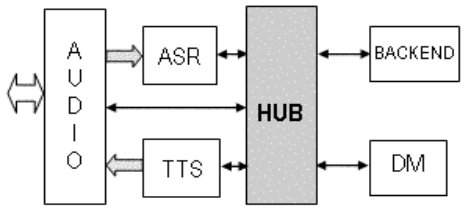

Fig. 3: IRKR communicator

The high effectiveness of dialog creating, clear XML syntax, unpretending maintainability and good portability of created voice services are further advantages of VoiceXML based DM.

Plug & play approach and high modularity are the advantages of Hub architecture.

5 CONCLUSION

The Galaxy architecture with VoiceXML based DM is a good resource to running dialogue services in a real application which are providing some information through the phone (for example call centers, information centers, weather information, bank services and others).

Our DM is not fully implemented yet. VoiceXML interpreter actually interprets VoiceXML 1.0 language. It will be upgraded to the version 2.0 in the next months.

The Weather information and Train information services are running on our Spoken Language Interactive System. They can be accessed on the phone number +42155 602 2297. For more information visit the IRKR web: http://irkr.tuke.sk

REFERENCES

[1] Juhár,J.-Čižmár,A.-Hintoš,Ľ: *Speech-enabled information services*. International Carpathian Control Conference ICCC2004, Zakopane, Poland, May,25-28,2004, pp. 837-842, 2004.

[2] Galaxy communicator website http://communicator.sourceforge.net/

[3] L. Boyer et al.: *Voice eXtensible markup Language 1.0.*, W3C NOTE, May 2000, http://www.w3.org

[4] Lihan,S.-Juhár,J.-Čižmár.A.: *Crosslingual and Bilingual Speech Recognition with Slovak and Czech SpeechDat-E Databases*. In Proc. of the 9th European Conference on Speech Communication and Technology Interspeech 2005, pp.225-228, Lisbon, Portugal, Sept 5.-8., 2005.(ISSN 1018-4074)